

Computed Tomography-based Radiomics for Risk Stratification in Prostate Cancer

Citation for published version (APA):

Osman, S. O. S., Leijenaar, R. T. H., Cole, A. J., Lyons, C. A., Hounsell, A. R., Prise, K. M., O'Sullivan, J. M., Lambin, P., McGarry, C. K., & Jain, S. (2019). Computed Tomography-based Radiomics for Risk Stratification in Prostate Cancer. *International Journal of Radiation Oncology Biology Physics*, 105(2), 448-456. <https://doi.org/10.1016/j.ijrobp.2019.06.2504>

Document status and date:

Published: 01/10/2019

DOI:

[10.1016/j.ijrobp.2019.06.2504](https://doi.org/10.1016/j.ijrobp.2019.06.2504)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Physics Contribution

Computed Tomography-based Radiomics for Risk Stratification in Prostate Cancer



Sarah O.S. Osman, PhD,^{*,†} Ralph T.H. Leijenaar, PhD,[‡]
Aidan J. Cole, PhD, FRCR,^{*,§} Ciara A. Lyons, MD,^{*,§}
Alan R. Hounsell, PhD,^{*,†} Kevin M. Prise, PhD,^{*}
Joe M. O'Sullivan, FRR (RCSI), PhD,^{*,§} Philippe Lambin, MD PhD,[‡]
Conor K. McGarry, PhD,^{*,†} and Suneil Jain, FRCR PhD^{*,§}

^{*}Centre of Cancer Research and Cell Biology, Queen's University Belfast, Belfast, United Kingdom;

[†]Radiotherapy Physics, Northern Ireland Cancer Centre, Belfast Health and Social Care Trust, Belfast, United Kingdom; [‡]The D-Lab: Decision Support for Precision Medicine, GROW - School for Oncology and Developmental Biology, Maastricht University Medical Centre, Maastricht, the Netherlands; and

[§]Clinical Oncology, Northern Ireland Cancer Centre, Belfast Health and Social Care Trust, Belfast, United Kingdom

Received Jan 15, 2019. Accepted for publication Jun 14, 2019.

Summary

The clinical behavior of localized prostate cancer is highly variable, with some cancers remaining indolent for many years and others progressing to metastatic disease in a short time. In this work, the value of computed tomography—based radiomics for risk stratification in prostate cancer was investigated. Our results show that radiomics features from routinely

Purpose: To explore the role of Computed tomography (CT)-based radiomics features in prostate cancer risk stratification.

Methods and Materials: The study population consisted of 506 patients with prostate cancer collected from a clinically annotated database. After applying exclusion criteria, 342 patients were included in the final analysis. CT-based radiomics features were extracted from planning CT scans for prostate gland—only structure, and machine learning was used to train models for Gleason score (GS) and risk group (RG) classifications. Repeated cross-validation was used. The discriminatory performance of the developed models was assessed using receiver operating characteristic area under the curve (AUC) analysis.

Results: Classifiers using CT-based radiomics features distinguished between $GS \leq 6$ versus $GS \geq 7$ with $AUC = 0.90$ and $GS 7(3 + 4)$ versus $GS 7(4 + 3)$ with $AUC = 0.98$. Developed classifiers also showed excellent performance in distinguishing low versus high RG ($AUC = 0.96$) and low versus intermediate RG ($AUC = 1.00$), but poorer performance was observed for $GS 7$ versus $GS > 7$ ($AUC = 0.69$). An overall

Corresponding author: Sarah O. S. Osman, PhD; E-mail: s.osman@qub.ac.uk

This work was supported by grants from Prostate Cancer UK, the Belfast-Manchester Movember Centre of Excellence (CE013-2-004), and the Northern Ireland Health and Social Care Trust R&D Office grant (COM/4965/14).

Disclosures: P.L. reports grants and sponsored research from Oncoradiomics, ptTheragnostic, and an advisor fee from Oncoradiomics; P.L. holds 2 patents for radiomics, claims 1 nonpatentable invention (software)

licensed to Oncoradiomics, and has (minority) shares in Oncoradiomics; R.T.H.L. reports personal fees and other from OncoRadiomics, outside the submitted work, and has a patent pending (EP3207521A1). K.M.P. reports personal fees from Bayer, outside the submitted work; J.M.O. reports grants and personal fees from Bayer, personal fees from Janssen, personal fees from Astellas, and personal fees from Sanofi, outside the submitted work.

Supplementary material for this article can be found at <https://doi.org/10.1016/j.ijrobp.2019.06.2504>.

acquired planning computed tomographic scans can provide insight into prostate cancer aggressiveness in a noninvasive manner.

modest performance was observed for validation on holdout data sets with the highest AUC of 0.75 for classifiers of low versus high RG and an AUC of 0.70 for GS 7 versus GS > 7.

Conclusions: Our results show that radiomics features from routinely acquired planning CT scans could provide insights into prostate cancer aggressiveness in a noninvasive manner. Assessing models on training data sets, the classifiers were especially accurate in discerning high-risk from low-risk patients and in classifying GS 7 versus GS > 7 and GS 7(3 + 4) versus GS 7(4 + 3); however, classifiers were less adept at distinguishing high RG versus intermediate RG. External validation and prospective studies are warranted to verify the presented findings. These findings could potentially guide targeted radiation therapy strategies in radical intent radiation therapy for prostate cancer. © 2019 Elsevier Inc. All rights reserved.

Introduction

Prostate cancer is one of the most frequently diagnosed cancers globally; it is the most common cancer in males in the United Kingdom, with approximately 130 new cases per day.¹ Other than skin cancer, prostate cancer is also the most common cancer in males in the United States, with approximately 64,690 new cases per year.² Traditionally, prostate cancer is stratified into risk categories using prostate-specific antigen (PSA) level, biopsy-based Gleason scores (GSs), and TNM staging.³ Although extremely useful in assessing prostate cancers, GSs are typically determined from a limited number of transrectal ultrasound (TRUS)—guided biopsies that are randomly sampled. This poses a challenge because prostate cancer is associated with significant intratumor heterogeneity.⁴ The behavior of prostate cancer is also highly variable, with some tumors progressing rapidly and others remaining indolent for many years. These factors make early diagnosis, accurate prognosis, and tumor risk stratification critical to increase the therapeutic ratio and to avoid unnecessary intervention.

Medical scans are conventionally used for diagnosis and treatment planning; however, in recent years, this role has expanded to include the extraction and quantification of imaging features (ie, radiomics analysis) to serve as imaging biomarkers for staging and to enhance the performance of prognostic and predictive models.⁵⁻¹⁰ The prognostic and predictive value of magnetic resonance imaging (MRI)—based radiomics features for prostate cancer has been demonstrated in a number of publications.⁸⁻¹¹ In recent years, there has been an increased focus on the development of multiparametric MRI (mpMRI) as a tool for effective image characterization, treatment planning, and response assessment.^{12,13} Despite its effectiveness, there is no clear consensus regarding the most important imaging biomarkers for prostate or their clinical applicability, possibly because of the lack of interoperability (ie, variable imaging protocols, scanners, software) or uncertainties in the manual definition of tumors.^{14,15} In contrast to MRI, planning computed tomography (CT) scans are typically highly standardized, and they are available for almost all patients treated with external beam

radiation therapy for outlining structures of interest, planning treatments, dose calculation, and as a reference for image guidance. Several studies have reported high repeatability and reproducibility of several CT-based radiomics features extracted from patient and from phantoms scans.¹⁶ Only a limited number of studies have investigated reproducibility with MRI scans reporting that MRI-based radiomics features are less robust.^{16,17} In radiation therapy, standardized planning CT scans are frequently assessed for CT number per electron density calibration, uniformity, and noise because this is critical for dose calculation; however, because MRI was traditionally used for visualization, less standardization has typically been used. Now that the role of medical imaging and radiomics analysis has been realized, there are more international efforts to standardize and control differences in medical scans to allow their optimal use for prospective imaging biomarkers studies.^{18,19}

To date, no studies have examined the role of non-contrast CT-based prostate radiomics in risk stratification or prognosis. This could be partially due to the poor visual discrimination of prostate cancer on CT scans; however, accumulating evidence of the value of imaging in diagnosis and prognosis of prostate cancer warrants more comprehensive investigation.⁵⁻¹¹

We hypothesize that the computerized interpretation of prostate intensity levels and their spatial distribution, as quantified from CT scans, will reveal additional information about prostate cancer, and radiomics analysis could serve as virtual biopsies.

In this study, we investigate the value of CT-based radiomics features to classify prostate cancer GS and risk group (RG). To our knowledge, this is the first investigation of noncontrast CT-based radiomics features for prostate cancer.

Methods and Materials

The use of CT scans and disease staging information of the study population in the present study was approved by the Belfast Health and Social Care Trust (BHSCT) with research governance permission number 18042CMcG-SS.

Study population

All patients with localized prostate cancer treated at the Northern Ireland Cancer Centre (NICC), between January 1, 2005, and December 31, 2009, were considered in this study. Patients were treated with 70 to 74 Gy external beam radiation therapy in 2-Gy fractions with 3-dimensional conformal or intensity modulated radiation therapy techniques over 7 to 7.5 weeks. Patients had short (≤ 6 months) or long-course (6–36 months) androgen-deprivation therapy commenced at least 3 months before radiation with luteinizing hormone-releasing hormone agonists or antiandrogens monotherapy (bicalutamide 150 mg daily dose). For patients in this cohort, GS was determined from TRUS biopsies. The median number of cores was 12 (range, 6–21), and the median volume of the tissue involved was 20% (range, 1%–90%; statistics available from 53% of patients). The median volume of tissue involved in different GS subgroups were: GS 6 = 7.5%; GS 7(3 + 4) = 18.8%; GS 7(4 + 3) = 20%; and GS > 7 = 20%. Patients were assigned to low-, intermediate-, and high-risk groups based on GS, initial PSA (iPSA), and TNM stage as per the United Kingdom National Institute for Health and Care Excellence guidelines.²⁰

CT image acquisition and image segmentation

Five hundred and six patients were initially identified for this study, of whom 447 had complete data sets (planning CT scans (pCT) and structure sets in digital imaging and communications in medicine (DICOM) format) to qualify for inclusion in this study. All patients were scanned at the NICC using our local pelvis protocol on one of two CT scanners: Emotion 6 (Siemens Health Care, Erlangen, Germany) and LightSpeed RT (GE Medical Systems, Chicago). Patients were scanned in helical mode using the following acquisition parameters: tube current, 370 to 437 mAs for the GE scanner and 130 to 194 mAs for Siemens; tube voltage, 120 kV(p) for GE and 130 kV(p) for Siemens; slice thickness, 2.5 mm; standard convolution kernel for GE and B41s for Siemens; 512 × 512 pixels; pixel spacing, 0.98 to 1.17 mm. All DICOM images and structure sets were exported into CERR software and converted to the MATLAB data file format.²¹ To reduce contouring variability, all prostate structures were reviewed and adjusted when necessary to construct a prostate gland-only structure (PO) by a single observer and reviewed by an experienced consultant clinical oncologist. The PO structure, which includes the whole prostate, was used for further analysis. Only CT scans with a 2.5 mm slice thickness and no artifacts were included in the analysis (eg, patients with prosthetic hips were excluded). Using these selection criteria, 342 scans were then identified as suitable for further analysis. Risk groupings are detailed in Tables 1 and 2 along with the number of patients assigned to each group.

Table 1 Patient grouping based on Gleason Score (GS)

Gleason Score	Number of patients
6 (3 + 3)	100
7 (3 + 4)	87
7 (4 + 3)	49
>7	106 (GS 8 = 55; GS 9 = 51)

Radiomics features extraction

RadiomiX software (Oncoradiomics, Liège, Belgium) was used to extract 1618 radiomic features from PO structures for each patient. After an initial investigation on optimal radiomics extraction settings (supplementary material Appendix E1; available online at <https://doi.org/10.1016/j.ijrobp.2019.06.2504>), CT scans were resampled to 2.5 mm isotropic voxels and the range of Hounsfield units (HU) was discretized to 10 HU bins before features extraction. Radiomics features extracted were the first-order gray-level statistics from the intensity level histograms, and features based on: gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), gray-level size-zone matrix (GLSZM), gray-level distance-zone matrix (GLDZM), neighborhood gray-tone difference matrix (NGTDM) and neighboring gray-level dependence matrix (NGLDM). The textural and statistics features were also calculated after applying 8 subbands of 3-dimensional wavelet transforms and 16 subbands of Laplacian of Gaussian (LoG) transform to the scans. A detailed explanation of the extracted features and their mathematical definitions are published elsewhere.^{6,22} In this analysis, geometric features (except for prostate volume) were excluded because the entire prostate gland was contoured and analyzed rather than tumor specific regions.

Test–retest cohort

Twenty patients with prostate cancer with available pre-treatment repeated pCT scans were also included in this study. These patients underwent repeated scans either for suboptimal bladder or rectal filling or for the temporary

Table 2 Patient grouping based on risk group

Risk group	UK NICE criteria	Number of patients
Low risk	T1 or T2a, Gleason score = 6, and PSA <10 ng/mL	n = 33
Intermediate risk	T2b or T2c, or Gleason score = 7, or PSA = 10–20 ng/mL	n = 76
High-risk	>T2c, Gleason score = 8–10, or PSA > 20 ng/mL	n = 233

presence of a rectal gas pocket on the first pCT scans. The average time between scans and rescans was 11 days (range, 0-28 days). Radiomics analysis of PO structures were conducted on the scans and the rescans. The intraclass correlation coefficient (ICC)^{23,24} was calculated to provide an indication of the test–retest repeatability and reproducibility of the features extracted. The intraclass correlation coefficient ranged between 0 and 1, with 0 indicating no reliability and 1 indicating perfect reliability. A cutoff value of 0.8 was used to arrive at a set of stable and robust features set for further analysis.

Feature selection and classification

For each developed model, stratified partitioning was used to randomly assign 80% of the data for model training and 20% for validation. To reduce the risk of models overfitting, the following feature reduction methods were used. The function *findCorrelation* in the R package *Caret* was used to remove (redundant) highly correlated features to reduce multicollinearity.²⁵⁻²⁷ This function searches through a correlation matrix of all the features and returns a vector of integers corresponding to columns to remove pair-wise correlations. Generalized linear models, via penalized maximum likelihood, were fitted using the R package *glmnet* and the default parameter tuning with the *Caret* interface was used. Two methods of variable selection were implemented: regularization through (1) the least absolute shrinkage and selection operator (LASSO) and (2) ElasticNet. Before the analysis, radiomics features were scaled and centered using *Caret*'s function *PreProc*. *PreProc* method “centre” centres feature values by subtracting the mean and the *PreProc* method “scale” divides by the standard deviation.²⁵ A grid search was used to choose optimal model hyperparameters that minimize cross-validation errors, and hyperparameters one standard error away from optimal values were then used. The average predictive performance of each model on the training data set was estimated using 10-fold cross-validation repeated 100 times. Sample augmentation using the Synthetic Minority Oversampling Technique (SMOTE) for class balancing in training data was also investigated.²⁸

Statistical analysis and model assessment

Following the guidelines of the transparent reporting of multivariable prediction model for individual prognosis or diagnosis statement for type 2a studies (ie, study using a single data set; data are randomly split into 2 groups: one to develop the prediction model and one to evaluate its predictive performance),²⁹ Mann-Whitney *U* tests (Wilcoxon rank sum test) were performed to test the null hypothesis that feature values have equal medians in different patient risk groups against the alternative that they are not.

In this work, all models were trained with and without sample augmentation. Model performances were assessed

on training and on testing data sets using area under the receiver operating characteristic curve (AUC), accuracy (Acc), sensitivity, and specificity and for comparison with previous work.³⁰ Youden index (YI)³¹ was also reported ($YI = \text{specificity} + \text{sensitivity} - 1$). For all models, the AUC 95% confidence interval (CI) was computed with 2000 stratified bootstrap replicates.

Results

A flow diagram of the study cohort, exclusion criteria, and prem modeling feature filtering is shown in Figure 1. After applying our exclusion criteria, 342 patients' CT scans were eligible for analysis. Forty PO structures from 20 patients were included in the test-retest analysis (Figs. 1 and 2A).

The performance of the trained models for classifying patients into different GS and risk groups was assessed in multiple steps (Tables 3 and 4). Sample augmentation was used to balance training data when required (detailed analysis of performance of models developed on non-augmented and augmented data are provided in Tables E1 and E2; available online at <https://doi.org/10.1016/j.ijrobp.2019.06.2504>). For example, the first radiomics classifier was trained to distinguish low from high GS ($GS \leq 6$ vs $GS \geq 7$). Figure 2B shows a circular bar plot of a subset of robust radiomics features ($n = 114$, unfiltered) in the two patient groups $GS \leq 6$ and $GS \geq 7$. The length of the bar represents the median value of each feature (scaled and centered), each patient group is represented with a different color. The stars in the circular bar plots indicate features that had significantly different median values in different groups. A comprehensive list of significant features is provided in Appendix E2 (available online at <https://doi.org/10.1016/j.ijrobp.2019.06.2504>).

Assessing the apparent performance of the trained models on training data, excellent performance was observed in classifying patients with low GS ($GS \leq 6$) from high GS ($GS > 6$) and low-risk from high-risk patients with $AUC = 0.90$ (95% CI, 0.86-0.95) and $AUC = 0.96$ (95% CI, 0.93-0.98), respectively (Fig. 3; Tables 3 and 4). Excellent performance was observed for classifiers of GS 7(3 + 4) versus GS 7(4 + 3) with an AUC of 0.98 (95% CI, 0.96-1.00) with high specificity and sensitivity. Developed classifiers had lower performance in distinguishing GS 7 from $GS > 7$ ($AUC = 0.69$; 95% CI, 0.61-0.76). The most important variables (features) in distinguishing GS and risk groups were found to be the wavelet transforms of the NGTDM coarseness (HLL, HLH, HHH) and contrast (HHH), the low gray-level small dependence emphasis LGSDE of the NGLDM (HHL, LLL, LHL, HLL), and the (LLL) GLSZM intensity nonuniformity normalized. All these features capture different aspects of texture heterogeneity.

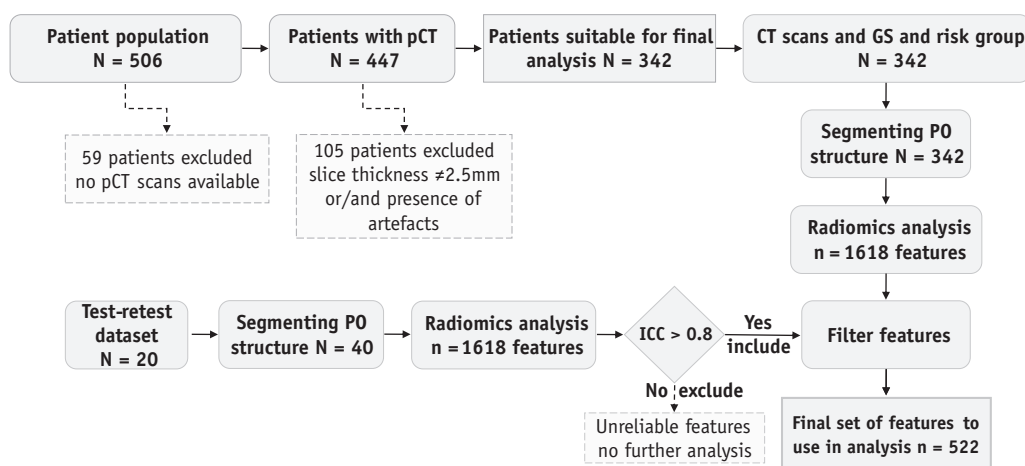


Fig. 1. Flow diagram of study cohort, radiomics workflow, and radiomics features generated. Patient numbers denoted by *N* and features by *n*.

All models were subsequently validated on unseen test data. The resulting AUC are presented in Tables 3 and 4 and Figure 3. The best observed performance was for classifiers of low-risk versus high-risk cancers, with an AUC of 0.75 (95% CI, 0.48-1.00) and for GS 7 versus GS > 7 (AUC = 0.70; 95% CI, 0.48-0.81). Lower performance was observed for models of GS 7(3 + 4) versus GS(4 + 3)

and low risk versus intermediate risk (AUC = 0.65 and 0.49, respectively).

In Figure 3, examples of AUC for different radiomics-based GS and RG classification are presented for models developed using ElasticNet on the training data and on testing data sets. Multiclass models for GS and RG classification were also explored and results are presented in

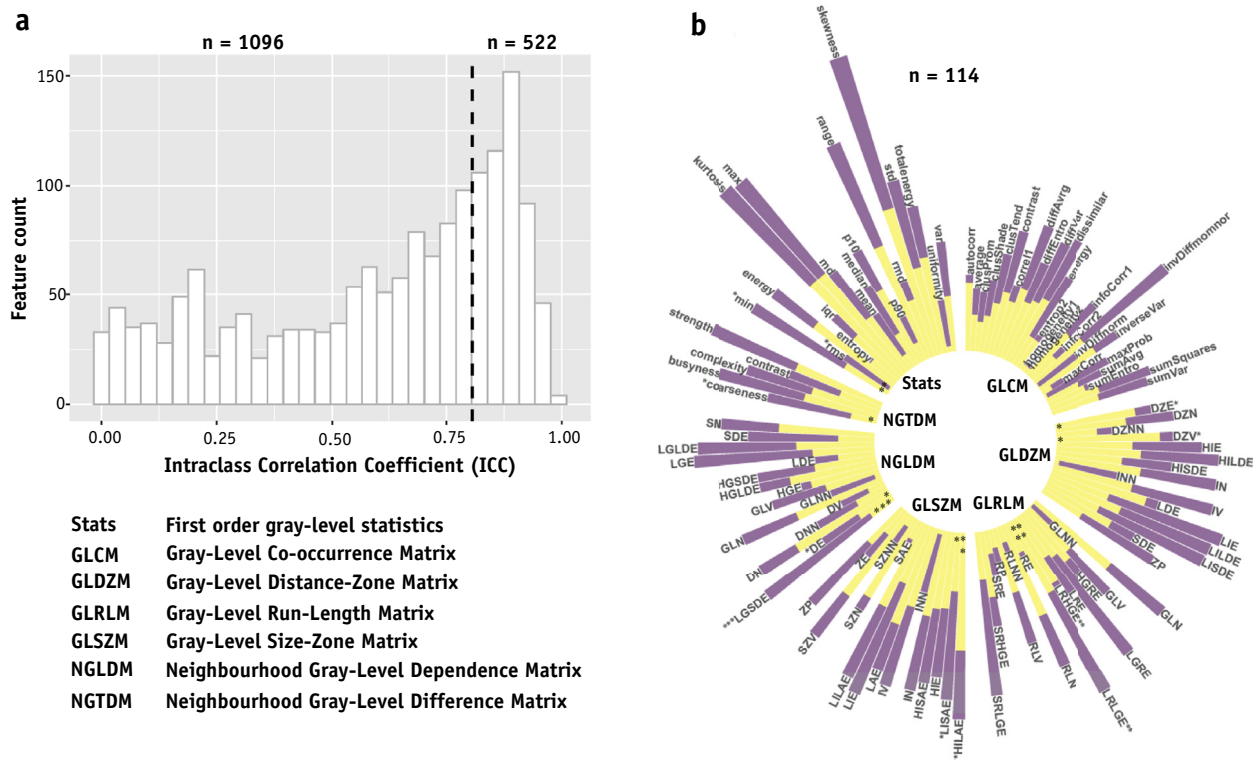


Fig. 2. (A) Histogram of the interclass correlation coefficients for radiomics features (*n* = 1618) from the test-retest data. The black dotted line indicates the cutoff value of 0.8 for reproducible features (*n* = 522). (B) A circular stacked bar plot of stable radiomics features (excluding filtered features) were scaled and centered for this presentation. Yellow and purple colors in the bar indicate median value of feature in patients with low and high GS cohorts, respectively. * indicates that a feature is significantly different between the two cohorts with Mann-Whitney *U* test **P* ≤ .05, ***P* ≤ .01, ****P* ≤ .001. (A color version of this figure is available at <https://doi.org/10.1016/j.ijrobp.2019.06.2504>.)

Table 3 Accuracy results for Gleason Score classification using Lasso and elastic Net regularization

Gleason Score	Training			Gleason Score	Testing		
	AUC	Accuracy	YI		AUC	Accuracy	YI
LASSO							
GS 6 vs > 6	0.77	0.65	0.50	GS 6 vs > 6	0.59	0.59	0.27
n = (80 vs 194)*	(0.73-0.84)	(0.59-0.71)		n = (19 vs 48)	(0.42-0.76)	(0.46-0.71)	
GS 7(3 + 4) vs 7(4 + 3)	0.97	0.86	0.66	GS 7(3 + 4) vs 7(4 + 3)	0.65	0.5	−0.13
n = (70 vs 40)	(0.95-1)	(0.79-0.92)		(17 vs 9)	(0.41-0.90)	(0.30-0.70)	
GS 7 vs > 7	0.67	0.62	0.13	GS 7 vs > 7	0.70	0.58	0.06
n = (109 vs 85)	(0.59-0.74)	(0.55-0.69)		(27 vs 21)	(0.54-0.86)	(0.43-0.72)	
ElasticNet							
GS 6 vs > 6	0.90	0.86	0.76	GS 6 vs >6	0.60	0.59	0.27
n = (80 vs 194)*	(0.86-0.95)	(0.81-0.90)		n = (19 vs 48)	(0.43-0.77)	(0.46-0.71)	
GS 7(3 + 4) vs 7(4 + 3)	0.98	0.92	0.80	GS 7(3 + 4) vs 7(4 + 3)	0.57	0.65	0.15
n = (70 vs 40)	(0.96-1.00)	(0.85-0.96)		n = (17 vs 9)	(0.30-0.84)	(0.44-0.83)	
GS 7 vs > 7	0.69	0.64	0.21	GS 7 vs > 7	0.65	0.58	0.07
n = (109 vs 85)	(0.61-0.76)	(0.57-0.71)		n = (27 vs 21)	(0.48-0.81)	(0.43-0.72)	

Abbreviations: AUC = area under the curve; YI = Youden index (Sensitivity + Specificity - 1); vs = versus; n = number of patients.

* n is the number of patients in each cohort. The asterisk indicates that training data were augmented before model training.

Table E3 (available online at <https://doi.org/10.1016/j.ijrobp.2019.06.2504>). Both GS and RG multiclass classifiers (built on SMOTEd data) had reasonable apparent performance on training data; however, only the RG multiclass classifier showed reasonable accuracy on the test data set (Acc = 0.70; 95% CI, 0.47-0.71). Applying sample augmentation through SMOTE, all the classifiers outperformed those built on nonaugmented samples with all models showing high sensitivity and specificity for all tested cases (**Appendix E3**; available online at <https://doi.org/10.1016/j.ijrobp.2019.06.2504>).

None of the developed classifiers had prostate volume in the final model. Pairwise *t* tests were conducted to test differences in prostate volume different risk groups. No

significant differences in prostate volume were found between different GS groups or between different androgen-deprivation therapy types; however, prostate volume was significantly larger in the intermediate- versus high-risk group (*P* = .02; **Figure E1**; available online at <https://doi.org/10.1016/j.ijrobp.2019.06.2504>).

Pearson and Spearman correlation coefficients were also calculated for the 20 most important variables, picked by each classifier (**Figure E2**; available online at <https://doi.org/10.1016/j.ijrobp.2019.06.2504>). The cross-correlation of patients' (initial) PSA, GS, risk group (RG), and prostate volumes (PV) with radiomics features revealed only weak correlations between individual radiomics features and clinical factors. It is also observed that several

Table 4 Accuracy results for prostate cancer risk group classification using Lasso and elasticNet regularization

Risk Group	Training			Risk Group	Testing		
	AUC	Accuracy	YI		AUC	Accuracy	YI
LASSO							
Low vs high n = (26 vs 187)*	0.96 (0.94-0.99)	0.89 (0.84-0.93)	0.87	Low vs high n = (6 vs 46)	0.71 (0.46-0.96)	0.79 (0.65-0.89)	−0.29
Low vs intermediate n = (33 vs 76)	1.00 (1.00-1.00)	1.00 (0.96-1.00)	1.00	Low vs intermediate n = (6 vs 15)	0.56 (0.21-0.90)	0.71 (0.48-0.89)	0.30
Intermediate vs high n = (61 of 187)*	1.00 (1.00-1.00)	0.99 (0.99-0.98)	0.97	Intermediate vs high n = (15 vs 46)	0.49 (0.33-0.63)	0.54 (0.41-0.67)	−0.10
ElasticNet							
Low vs high n = (26 of 187)*	0.96 (0.93-0.98)	0.88 (0.65-0.89)	0.87	Low vs high n = (6 vs 46)	0.75 (0.48-1.00)	0.83 (0.72-0.93)	0.54
Low vs intermediate n = (33 of 76)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00	Low vs intermediate n = (6 vs 15)	0.56 (0.21-0.91)	0.71 (0.48-0.89)	0.30
Intermediate vs high n = (61 of 187)*	1.00 (1.00-1.00)	0.99 (0.99-0.98)	0.97	Intermediate vs high n = (15 vs 46)	0.49 (0.33-0.63)	0.56 (0.43-0.68)	−0.40

Abbreviations: AUC = area under the curve; YI = Youden index (Sensitivity + Specificity - 1); vs = versus; n = number of patients..

* n is the number of patients in each cohort. The asterisk indicates that training data were augmented before model training.

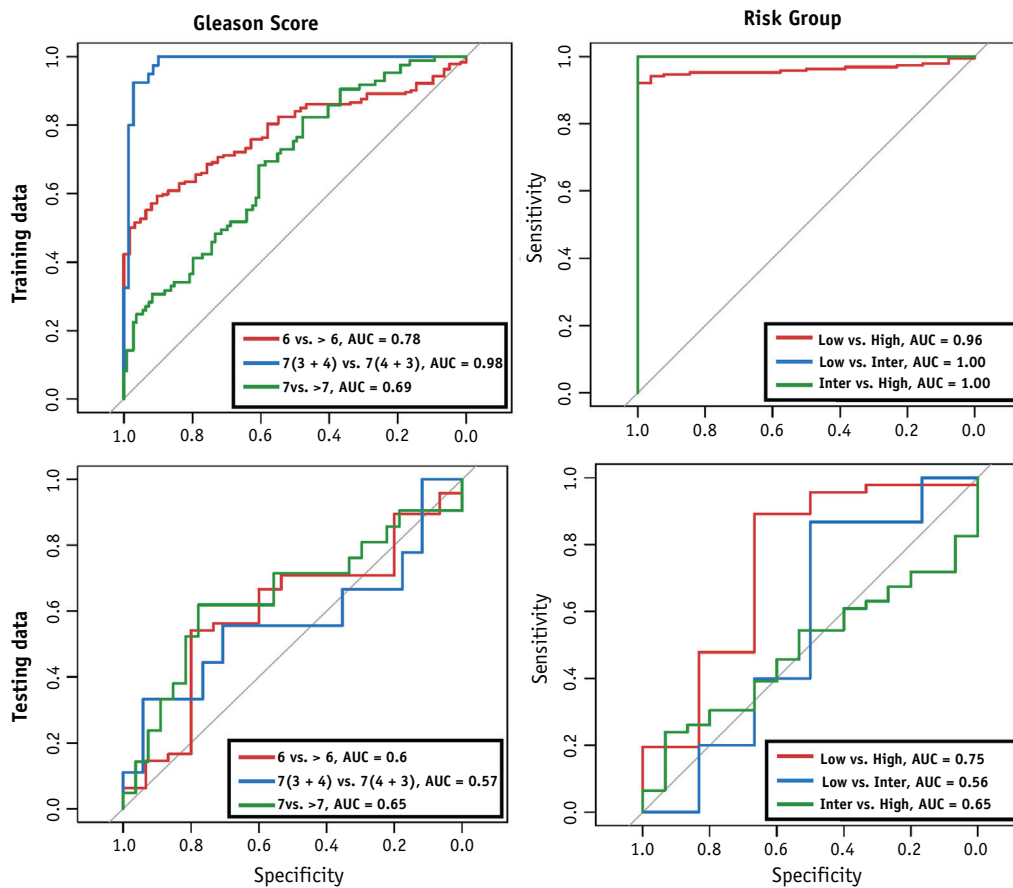


Fig. 3. The receiver operating characteristics for Gleason score (left) and risk group (right) classification with ElasticNet. Classifiers performance on training (top) and testing data (bottom) is presented.

radiomics features exhibit monotonic relationships indicated by the high values for their Spearman correlation coefficients. A Venn diagram showing the intersections of the top features selected by LASSO for different classification problems is presented in Figure E3 (available online at <https://doi.org/10.1016/j.ijrobp.2019.06.2504>).

Discussion

In this study, the value of CT-based radiomics analysis in prostate GS and risk group classification is demonstrated. This work describes the extraction of quantitative imaging features and the application of machine learning methods for features selection and reduction and model training. A test–retest analysis was conducted to investigate the robustness and stability of extracted features and to eliminate nonreliable features. Given the retrospective nature of the database used in this study and limitations associated with the variable scan–rescan data (scanners, scan parameters, and time gap between scans), this test–retest analysis is likely to eliminate more features than truly required. However, this method was adopted to provide a conservative measure to account for the variations in the data set used in this analysis.

Previous studies on prostate cancer radiomics mainly focused on the use of mpMRI.^{8–11,30,32–38} Several studies have demonstrated the value of mpMRI-based features for prostate cancer detection,^{8,32–34} prostate cancer aggressiveness and GS classification,^{30,35} biochemical failure risk prediction,^{36,37} and in targeted radiation therapy planning.³⁸

Fehr et al³⁰ investigated the use of MRI-based texture features in cancer detection and GS assessment in prostate cancer patients. They showed that accurate models for GS classification could be achieved by using machine-learning methods to combine several apparent diffusion coefficients ADC and T2 measures, even in highly imbalanced data sets. Similar to our work, they also showed that using sample augmentation methods always resulted in better performing classification models in all the cases.

Only one study was found for perfusion-CT–based radiomics for prostate cancer.⁷ Motivated by previous studies showing the correlation of perfusion-CT parameters with PSA values, GS, and TNM stage,^{39,40} Tanasini-Lang et al⁷ investigated perfusion-CT–based radiomics for GS classifications. In their study, before radical prostatectomy, all patients ($n = 41$) had perfusion CT, and perfusion parameters calculated from PO contours. Three perfusion maps were calculated (ie, blood volume, blood flow (BF),

mean transit time), and 1701 radiomics features from these maps were calculated for each patient. Using principle component analysis, 10 feature groups were identified. From each group, a single radiomics feature (that correlated best with the group) was chosen to represent the group in a multivariate regression analysis to predict the postsurgical GS. They found one radiomics feature, the HHL wavelet transform of BF joint average, to be prognostic for GS (7 vs >7). Comparing patients with GS 7(3 + 4) versus GS 7(4 + 3), both BF HHH fractal dimension and blood volume HLH root mean square were found different between the two groups. Interestingly, they also reported that for 28% of patients there was a difference in grading between the biopsy and postsurgical histopathology.⁷

Although it is not possible to correlate the CT-based radiomics features in this study directly with previous work on MRI or perfusion-CT, a general comparison of the performance of final classification models is presented for nonaugmented data. CT-based radiomics ElasticNet-based classifiers of GS 6 versus GS > 6 ($n = 80$ vs 194) achieved an accuracy of 0.72 with $YI = 0.04$ compared with 0.83 accuracy and $YI = 0.03$ reported for the MRI-based classifier by Fehr et al³⁰ ($n = 34$ vs 159). Excellent performance was observed in this work for classifiers of GS 7(3 + 4) versus GS 7(4 + 3) with $Acc = 0.98$ and $YI = 0.80$ ($n = 70$ vs 40) compared with $Acc = 0.83$ and $YI = 0.11$ in Gamer et al²⁴ ($n = 114$ vs 26). CT-based classifier for GS (7 vs >7) showed lower performance ($AUC = 0.69$; $Acc = 0.64$) compared with perfusion-CT based models ($AUC = 0.81$)⁷; however, the number of patients in Tanadini-Lang et al⁷ was much smaller ($n = 32$ vs 8) than the number of patients in the present study ($n = 109$ vs 85).

In this study, the role of radiomics features in classifying patients with prostate cancer into different risk groups was also investigated. Despite the highly imbalanced data sets, excellent apparent performance was observed for classifiers of low versus high-risk group ($n = 26$ vs. 187) groups ($Acc = 0.98$; $AUC = 1.00$; $YI = 0.81$) and low- versus intermediate-risk groups ($Acc = 1.00$; $AUC = 1.00$; $YI = 1.00$). However, poor performance was observed for classifiers of intermediate versus high risk ($Acc = 0.75$; $AUC = 0.68$; $YI = 0$), Table E2 (available online at <https://doi.org/10.1016/j.ijrobp.2019.06.2504>).

In all the studies discussed thus far, the data were used to develop the models, and the performance of classifiers was assessed using resampling techniques. In this study, classifiers were also tested on unseen data showing overall modest performance. This result could be influenced by the small number of patients in test data sets and by the class imbalances. Nonetheless, this study shows that CT-based radiomics features can provide insight into the nature of prostate cancer beyond the scope of visual assessment. Our models classified patients with low SG/RG versus high GS/RG and with GS 7(3 + 4) versus GS 7(4 + 3) with high accuracy. The developed classifiers were especially accurate in identifying high-risk patients with a high Gleason score. This study does not suggest replacing MRI-based

radiomics for prostate cancer risk assessment. This study demonstrates that CT-based radiomics carry information that can be used to improve current risk stratification methods. Combining CT- and MRI-based radiomics could give more accurate assessments of prostate cancer aggressiveness, and this subject needs further investigation. It is acknowledged that this study has some limitations. First, this is a retrospective study with no external validation data set. In an ongoing study, three data sets for external validation are being prepared; the first data set is from our own institution sampled from a later period (2010-2015), and the two other cohorts will come from other external institutions. Moreover, radiomics features in this study were extracted from the PO structures and not prostate cancer explicitly. Although this is a more accurate representation of current prostate cancer RT treatments, in several investigations on MRI-based prostate cancer, radiomics features were found to be different in cancers arising in the transitional zone from those in the peripheral zone of the prostate.^{30,32} This limitation can be addressed in future prospective studies by contouring suspicious areas of the prostate on CT scans aided by MRI fusion.¹¹ This technique is used in clinical practice for defining boost volumes of dominant intraprostatic lesions. Cancer-specific contouring might also allow better understanding of the shape features. Moreover, the endpoint in this study was GS based on TRUS biopsies. It is expected that mpMRI guided biopsies will provide more accurate grading of prostate cancers and consequently allow the training of more accurate image-based classifiers. Another observation in this study was the correlation between prostate volume and some of the radiomics features included in the final models. As discussed by Welch et al⁴¹ on their work on radiomics safeguards, this might have implications for models' performance and generalizability. Therefore, these correlations between radiomics features and prostate volume should be investigated thoroughly in future studies.

Despite these limitations, the current study is proof of concept that standardly obtained pCT scans can reproducibly characterize risk stratification in prostate cancer and can guide future studies investigating bespoke radiation therapy based on CT radiomics. External validation and prospective studies are warranted to verify our findings.

References

1. Prostate cancer incidence statistics. Available at: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer/incidence>. Accessed September 14, 2018.
2. Key statistics for prostate cancer. Available: <https://www.cancer.org/cancer/prostate-cancer/about/key-statistics.html>. Accessed September 14, 2018.
3. D'Amico AV, Whittington R, Malkowicz SB, et al. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA* 1998;280:969-974.

4. Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 2012;366:883-892.
5. Coates J, El Naqa I. Outcome modeling techniques for prostate cancer radiotherapy: Data, models, and validation. *Phys Medica* 2016;32:512-520.
6. Lambin I, Zindler J, Vanneste BG, et al. Decision support systems for personalized and participative radiation oncology. *Adv Drug Deliv Rev* 2017;109:131-153.
7. Tanadini-Lang S, Bogowicz M, Veit-Haibach P, et al. Exploratory radiomics in computed tomography perfusion of prostate cancer. *Anticancer Res* 2018;38:685-690.
8. Algohary A, Viswanath S, Shiradkar R, et al. Radiomic features on MRI enable risk categorization of prostate cancer patients on active surveillance: Preliminary findings. *J Magn Reson Imaging* 2018;48:818-828.
9. Wang J, Wu CJ, Bao ML, Zhang J, Wang XN, Zhang Y. Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer. *Eur Radiol* 2017;27:4082-4090.
10. Cameron A, Khalvati F, Haider MA, Wong A. MAPS: A quantitative radiomics approach for prostate cancer detection. *IEEE Trans Biomed Eng* 2016;63:1145-1156.
11. Ahmed HU, El-Shater Bosaily A, Brown LC, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): A paired validating confirmatory study. *Lancet* 2017;389:815-822.
12. Schmidt MA, Payne GS. Radiotherapy planning using MRI. *Phys Med Biol* 2015;60:R323-R361.
13. Kerkmeijer LGW, Maspero M, Meijer GJ, van der Voort van Zyp JRN, de Boer HCJ, van den Berg CAT. Magnetic resonance imaging only workflow for radiotherapy simulation and planning in prostate cancer. *Clin Oncol* 2018;30:692-701.
14. Jambor I. Optimization of prostate MRI acquisition and post-processing protocol: A pictorial review with access to acquisition protocols. *Acta Radiol Open* 2017;6: 2058460117745574.
15. Sonn GA, Fan RE, Ghanouni P, et al. Prostate magnetic resonance imaging interpretation varies substantially across radiologists. *Eur Urol Focus* 2017. <https://doi.org/10.1016/j.euf.2017.11.010> [e-pub ahead of print]. Accessed September 14, 2018.
16. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and reproducibility of radiomic features: A systematic review. *Int J Radiat Oncol Biol Phys* 2018;102:1143-1158.
17. Baeßler B, Weiss K, Dos Santos DP. Robustness and reproducibility of radiomics in magnetic resonance imaging: A phantom study. *Invest Radiol* 2019;54:221-228.
18. HATT M, Vallières M, Visvikis D, A Zwanenburg. IBSI: an international community radiomics standardization initiative. *J Nucl Med* 2018;59, supplement 1 287.
19. A. Zwanenburg, S Leger, M Vallières, S Löck ,Image biomarker standardisation initiative. 2019; arXiv:1612.07003v9
20. Prostate cancer: diagnosis and management. Available at: <https://www.nice.org.uk/guidance/cg175/evidence/full-guideline-191710765>. Accessed September 14, 2018.
21. Deasy JO, Blanco AI, Clark VH. CERR: A computational environment for radiotherapy research. *Med Phys* 2003;30:979-985.
22. Aerts HJWL, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5.
23. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. *Psychol Rep* 1966;19:3-11.
24. Gamer M, Lemon J, Fellows I, Singh P. Irr package for R, version 084: 22-23; Available at <https://CRAN.R-project.org/package=irr>. Accessed September 14, 2018.
25. Kuhn M. Building predictive models in R using the caret package. *J Stat Software Artic* 2008;28:1-26.
26. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010;33:1-22.
27. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent. *J Stat Softw* 2011;39:1-13.
28. Torgo L. Data Mining with R: Learning with Case Studies. Boca Raton, FL: Chapman and Hall/CRC; 2010.
29. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *Eur Urol* 2015 ;67:1142-1151.
30. Fehr D, Veeraraghavan H, Wibmer A, et al. Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images. *Proc Natl Acad Sci* 2015;112:E6265-E6273.
31. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32-35.
32. Ginsburg SB, Algohary A, Pahwa S, et al. Radiomic features for prostate cancer detection on MRI differ between the transition and peripheral zones: Preliminary findings from a multi-institutional study. *J Magn Reson Imaging* 2017;46:184-193.
33. Chung AG, Shafiee MJ, Kumar D, et al. Discovery radiomics for multi-parametric MRI prostate cancer detection. *arXiv* 2015. 1509.00111v3.
34. Sidhu HS, Benigno Salvatore, Ganeshan Balaji, et al. Textural analysis of multiparametric MRI detects transition zone prostate cancer. *Eur Radiol* 2017;27:2348-2358.
35. Wibmer A, Hricak H, Gondo T, et al. Haralick texture analysis of prostate MRI: Utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores. *Eur Radiol* 2015;25:2840-2850.
36. Khémara G, Fargeas A, Gutiérrez-Carvajal RE, et al. Haralick textural features on T2-weighted MRI are associated with biochemical recurrence following radiotherapy for peripheral zone prostate cancer. *J Magn Reson Imaging* 2017;45:103-117.
37. Rakesh S, Ghose S, Jambor I, et al. Radiomic features from pre-treatment biparametric MRI predict prostate cancer biochemical recurrence: Preliminary findings. *J Magn Reson Imaging* 2018;Dec; 48:1626-1636.
38. Shiradkar R, Podder TK, Algohary A, Viswanath S, Ellis RJ, Madabhushi A. Radiomics based targeted radiotherapy planning (Rad-TRaP): A computational framework for prostate cancer treatment planning with MRI. *Radiat Oncol* 2016;11:148.
39. Stelmach A. Correlation between CT perfusion and clinico-pathological features in prostate cancer: A prospective study. *Med Sci Monit* 2015;21:153-162.
40. Huellner MW, Pauli C, Mattei A, et al. Assessment of prostate cancer with dynamic contrast-enhanced computed tomography using an en bloc approach. *Invest Radiol* 2014;49:571-578.
41. Welch ML, McIntosh C, Haibe-Kains B, et al. Vulnerabilities of radiomic signature development: The need for safeguards. *Radiother Oncol* 2019;130:2-9.